CLAIMS:

1. A method of selecting one or more variables for use with a statistical model, the method comprising the steps of:

creating a plurality of unique subsets of variables of multivariate data;

determining the performance of a discriminant rule when used with each of the subsets, the discriminant rule being based on multivariate normal class densities each having substantially diagonal covariance matrices; and

selecting the one or more variables from at least one of the subsets that result in a desired performance of the discriminant rule.

2. The method as claimed in claim 1, wherein the step of creating the plurality of unique subsets comprises the step of identifying a variable in the multivariate data that is not a member of a set of variables, and adding the identified variable to the set.

3. The method as claimed in any one of claims 1 or 2, wherein the step of determining the performance of the discriminant rule comprises assessing a prediction error rate of the discriminant rule.

4. The method as claimed in claim 3, wherein the prediction error rate is a cross-validated error rate.

5. The method as claimed in any one of the preceding claims, wherein the desired performance of the discriminant rule comprises the lowest possible prediction error rate of the discriminant rule.

6. The method as claimed in any one of the preceding claims, wherein the multivariate data comprises gene expression data.

7. Computer software which, when executed by a computer, enables the computer to carry out the steps defined in any one of the preceding steps.

8. A computer storage medium containing the software defined in claim 7.

9. A statistical model for predicting a class of an observation, wherein the model includes one or more variables that have been selected using the method defined in any one of claims 1 - 6.

10. An apparatus for selecting one or more variables for use with a statistical model, the system comprising:

data creating means arranged to create a plurality of unique subsets of variables of multivariate data;

a processing means arranged to determine the performance of a discriminant rule when used with each of the subsets, the discriminant rule being based on multivariate normal class densities each having substantially diagonal covariance matrices; and

a selecting means arranged to select the one or more variables from at least one of the subsets that results in a desired performance of the discriminant rule.

11. The apparatus as claimed in claim 10, wherein the data creating means is arranged to create the plurality of unique subsets by identifying a variable in the multivariate data that is not a member of a set of variables, and adding the identified variable to the set.

12. The apparatus as claimed in any one of claims 10 or 11, wherein the determining means is arranged to determine the performance of the discriminant rule by assessing a prediction error rate of the discriminant rule.

13. The apparatus as claimed in claim 12, wherein the prediction error rate is a cross-validated error rate.

14. The apparatus as claimed in any one of the preceding claims, wherein the desired performance of the discriminant rule comprises the lowest possible prediction error rate of the discriminant rule.

15. The apparatus as claimed in any one of claims 10 - 14, wherein the multivariate data comprises gene expression data.

16. The apparatus as claimed in any one of claims 10 - 15, wherein the data creating means, processing means and selecting means are in the form of a computer running software.